

## CLAIMS

1. A polypeptide comprising a variant CDRH3 region, wherein the CDRH3 region

5 comprises:

a) at least one structural amino acid position, wherein said structural amino acid position has a variant amino acid , wherein the variant amino acid is an amino acid found at that position in a randomly generated CDRH3 population at a frequency of at least one standard deviation above the average frequency for any amino acid at that position; and

b) at least one non-structural position, wherein the non-structural position has a variant amino acid.

2. A variable domain of a monobody comprising a variant CDRH3 region, wherein

15 the variant CDRH3 region comprises:

a) at least one structural amino acid position, wherein said structural amino acid position has a variant amino acid , wherein the variant amino acid is an amino acid found at that position in a randomly generated population at a frequency of at least one standard deviation above the average frequency for any amino acid at that position; and

b) at least one non-structural position, wherein the non-structural position has a variant amino acid.

3. The polypeptide according to claim 1 or 2, wherein the polypeptide is an antibody

25 variable domain of the Vh3 subgroup.

4. The polypeptide according to claims 1 or 2, wherein said at least one non-

structural position is a contiguous amino acid sequence of about 1 to 20 amino acids.

5. The polypeptide according to claims 1 or 2, wherein said at least one structural amino acid position is one or both the first two amino acid positions at the N-terminus of a heavy chain CDRH3.

5 6. The polypeptide according to claims 1 or 2, wherein said at least one structural amino acid position is at least one of the last 6 amino acids at the C-terminus of a heavy chain CDRH3.

7. The polypeptide according to claim 5, wherein the first N-terminal amino acid 10 position has a variant amino acid is selected from the group consisting of R, L, and V.

8. The polypeptide according to claim 5, wherein said at least one structural amino acid position is at position 96 in the CDRH3 of SEQ ID NO:137 and the variant amino acid is R, L, or V.

15  
9. The polypeptide according to claim 5, wherein the first amino acid position at the N-terminus has a variant amino acid selected from the group consisting of R, L and V, and the second amino acid position at the N-terminus is selected from the group consisting of I and L.

20  
10. The polypeptide according to claim 6, wherein said at least one structural amino acid position is a third and/or fourth amino acid position from the C-terminus.

25  
11. The polypeptide according to claim 10, wherein the fourth amino acid position from the C-terminus has a variant amino acid selected from the group consisting of M, R, G and W and the third amino acid position from the C-terminus has a variant amino acid selected from the group consisting of P, L, or V .

30  
12. The polypeptide according to claim 6, wherein the at least one structural amino acid position is selected from the amino acid position 100g, 100h, 100i, 100j, 101, 102 of SEQ ID NO:137 and mixtures thereof.

13. The polypeptide according to claim 12, wherein the variant amino acid at position 100i is amino acid selected from the group consisting of M, R, G and W; and wherein the variant amino acid at position 100j is amino acid selected from the group consisting of P, 5 L, or V; and wherein the variant amino at position 100g is amino acid selected from the group consisting of E, F and W.

14. The polypeptide according to claim 1 or 2, wherein said at least one non-structural position has a variant amino acid that is any one of the naturally occurring amino acids.

10

15. The polypeptide according to claim 1 or 2, wherein said at least one non-structural position has a variant amino acid encoded by a non-random codon set.

16. The polypeptide according to claims 1 or 2, wherein the said at least one structural amino acid position is the first two N-terminal amino acid positions, and the third and fourth positions from the C-terminus of the CDRH3 region.

15

17. The polypeptide according to claim 16, wherein the first two N-terminus amino acid positions are positions 96 and 97, and the third and fourth position from the C-terminus are 100i and 100j of SEQ ID NO:137.

20

18. The variable domain according to claim 2, wherein amino acid position 37 of the framework 2 region is a hydrophobic amino acid.

25

19. The variable domain according to claim 18, wherein amino acid position 37 is phenylalanine or tryptophan.

20

20. The variable domain according to claim 18, wherein the amino acid position 45 of framework 2 is selected from the group consisting of arginine, tryptophan, phenylalanine 30 and leucine.

21. A variable domain of claim 2, further comprising a heavy chain framework 3 region, wherein the amino acid position 91 of the framework 3 region is a phenylalanine tyrosine, or threonine.

5 22. The polypeptide of any of claims 1-21 which is a fusion polypeptide.

23. The polypeptide of claim 22 which is a fusion polypeptide fused to at least a portion of a viral coat protein.

10 24. The polypeptide of claim 23, wherein the viral coat protein is selected from the group consisting of p111, pv111, Soc, Hoc, 9pD, pV1 and variants thereof.

25. A polynucleotide molecule encoding a polypeptide of any of claims 1-24.

15 26. A replicable expression vector comprising a polynucleotide molecule of claim 25.

27. A host cell comprising the vector of claim 26.

28. A library comprising a plurality of vectors of claim 26, wherein the plurality of 20 vectors encode a plurality of polypeptides.

29. A polypeptide comprising a variant CDRH3 region, wherein the CDRH3 region comprises:

25      a) a N terminal portion that comprises at least one structural amino acid position, wherein said structural amino acid position has a variant amino acid , wherein the variant amino acid is an amino acid found at that position in a randomly generated CDRH3 population at a frequency of at least one standard deviation above the average frequency for any amino acid at that position;

30      b) a central portion that comprises at least one non-structural position, wherein the non-structural position has a variant amino acid; and

5                   c)a C-terminal portion that comprises at least one structural amino acid position, wherein said structural amino acid position has a variant amino acid , wherein the variant amino acid is an amino acid found at that position in a randomly generated CDRH3 population at a frequency of at least one standard deviation above the average frequency for any amino acid at that position.

30.       The polypeptide according to claim 29, wherein the polypeptide is a heavy chain variable domain of a monobody.

10      31.       The polypeptide according to claims 29, wherein said at least one non-structural position is a contiguous amino acid sequence of about 1 to 17 amino acids.

15      32.       The polypeptide according to claims 29, wherein said at least one structural amino acid position is one or both the first two amino acid positions at the N-terminus of a heavy chain CDRH3.

20      33.       The polypeptide according to claims 29, wherein said at least one structural amino acid position is at least one of the last 6 amino acids at the C-terminus of a heavy chain CDRH3.

25      34.       The polypeptide according to claim 32, wherein the first N-terminal amino acid position has a variant amino acid is selected from the group consisting of R, L, and V.

30      35.       The polypeptide according to claim 32, wherein said at least one structural amino acid position is at position 96 in the CDRH3 of SEQ ID NO:137 and the variant amino acid is R, L, or V.

35      36.       The polypeptide according to claim 32, wherein the first amino acid position at the N-terminus has a variant amino acid selected from the group consisting of R, L and V, and the second amino acid position at the N-terminus is selected from the group consisting of I and L.

37. The polypeptide according to claim 32, wherein the N terminal portion is no more than 4 amino acids.

5 38. The polypeptide according to claim 33, wherein said at least one structural amino acid position is a third and/or fourth amino acid position from the C-terminus.

10 39. The polypeptide according to claim 38, wherein the fourth amino acid position from the C-terminus has a variant amino acid selected from the group consisting of M, R, G and W and the third amino acid position from the C-terminus has a variant amino acid selected from the group consisting of P, L, or V .

15 40. The polypeptide according to claim 33, wherein the at least one structural amino acid position is selected from the amino acid position 100g, 100h, 100i, 100j, 101, 102 of SEQ ID NO:137 and mixtures thereof.

20 41. The polypeptide according to claim 40, wherein the variant amino acid at position 100i is amino acid selected from the group consisting of M, R, G and W; and wherein the variant amino acid at position 100j is amino acid selected from the group consisting of P, L, V, and W; and wherein the variant amino acid at position 100g is amino acid selected from the group consisting of E, F and W.

25 42. The polypeptide according to claim 33, wherein the C-terminal portion is not more than 6 amino acids.

43. The polypeptide according to claim 29, wherein said at least one non-structural position has a variant amino acid that is any one of the naturally occurring amino acids.

30 44. The polypeptide according to claim 29, wherein said at least one non-structural position has a variant amino acid encoded by a non-random codon set.

45. The polypeptide according to claim 29, wherein the central portion is no more than 20 amino acids.

5

46. The polypeptide according to claims 29, wherein the said at least one structural amino acid position is the first two N-terminal amino acid positions, and the third and fourth positions from the C-terminus of the CDRH3 region.

10 47. The polypeptide according to claim 46, wherein the first two N-terminus amino acid positions are positions 96 and 97, and the third and fourth position from the C-terminus are 100i and 100j of SEQ ID NO:137.

15 48. The variable domain according to claim 30, wherein amino acid position 37 of the framework 2 region is a hydrophobic amino acid.

49. The variable domain according to claim 48, wherein amino acid position 37 is phenylalanine or tryptophan.

20 50. The variable domain according to claim 30, wherein the amino acid position 45 of framework 2 is selected from the group consisting of arginine, tryptophan, phenylalanine and leucine.

25 51. A variable domain of claim 30, wherein the amino acid position 91 of the framework 3 region is a phenylalanine, tyrosine or threonine.

52. The polypeptide of any of claims 29-51 which is a fusion polypeptide.

30 53. The polypeptide of claim 52 which is a fusion polypeptide fused to at least a portion of a viral coat protein.

54. The polypeptide of claim 53, wherein the viral coat protein is selected from the group consisting of p111, pv111, Soc, Hoc, 9pD, pV1 and variants thereof.

55. A polynucleotide molecule encoding a polypeptide of any of claims 29-54.

5

56. A replicable expression vector comprising a polynucleotide molecule of claim 55.

57. A host cell comprising the vector of claim 56.

10

58. A library comprising a plurality of vectors of claim 56, wherein the plurality of vectors encode a plurality of polypeptides.

59. A polypeptide comprising a CDRH3, wherein the CDRH3 comprises an amino acid sequence having the formula of  $A_1-A_2-(A_3)_n-A_4-A_5$ ; wherein

15  $A_1$  is an amino acid selected from the group consisting of R, L, V, F, W and K;

$A_2$  is an amino acid selected from the group consisting of I, L, V, R, W and S;

$A_3$  is any naturally occurring amino acid and n can be 1-17;

20  $A_4$  is an amino acid selected from the group consisting of W, G, R, M, S, A and H;

$A_5$  is an amino acid selected from the group consisting of V, L, P, G, S, E and W.

60. A polypeptide comprising a CDRH3, wherein the CDRH3 comprises an amino acid sequence having the formula of  $A_1-A_2-(A_3)_n-A_4-A_5-A_6-A_7$ ; wherein

25  $A_1$  is an amino acid selected from the group consisting of R, L, V, F, W and K;

$A_2$  is an amino acid selected from the group consisting of I, L, V, R, W and S;

$A_3$  is any naturally occurring amino acid and n can be 1-17;

$A_4$  is an amino acid selected from the group consisting of W, G, R, M, S, A and H;

30  $A_5$  is an amino acid selected from the group consisting of V, L, P, G, S, E and W; and

A<sub>6</sub> and A<sub>7</sub> are any naturally occurring amino acid.

61. A variable domain of a monobody comprising a CDRH3 , wherein the CDRH3 comprises an amino acid sequence having the formula of A<sub>1</sub>—A<sub>2</sub>—(A<sub>3</sub>)<sub>n</sub>—A<sub>4</sub>—A<sub>5</sub>—A<sub>6</sub>—A<sub>7</sub>;

5 wherein

A<sub>1</sub> is an amino acid selected from the group consisting of R, L, V, F, W and K;

A<sub>2</sub> is an amino acid selected from the group consisting of I, L, V, R, W and S;

A<sub>3</sub> is any naturally occurring amino acid and n can be 1-17;

A<sub>4</sub> is an amino acid selected from the group consisting of W, G, R, M, S, A and

10 H;

A<sub>5</sub> is an amino acid selected from the group consisting of V, L, P, G, S, E and W;  
and

A<sub>6</sub> and A<sub>7</sub> are any naturally occurring amino acid.

15 62. The polypeptide according to claims 59 to 61, wherein

A<sub>1</sub> is R;

A<sub>2</sub> is I;

A<sub>4</sub> is W;

A<sub>5</sub> is V; and

20 n=11.

63. The polypeptide according to claims 59 to 61, wherein

A<sub>1</sub> is L;

A<sub>2</sub> is L;

25 A<sub>5</sub> is L; and

n=11.

64. The polypeptide according to claims 59 to 61, wherein

A<sub>1</sub> is V;

30 A<sub>2</sub> is L;

A<sub>4</sub> is R;

A<sub>5</sub> is V; and

n=11.

65. The polypeptide according to claims 59 to 61, wherein

5 A<sub>1</sub> is R;

A<sub>2</sub> is L; and

n=11.

10 66. The polypeptide according to claims 59 to 61, wherein n is 9 to 12.

67. The polypeptide according to claims 59 to 61, wherein amino acid or amino acids of A<sub>3</sub> are any one of the naturally occurring amino acids.

15 68. The polypeptide according to claims 59 to 61, wherein the amino acid or amino acids of A<sub>3</sub> are encoded by a nonrandom codon set.

69. A polypeptide comprising a CDRH3, wherein the CDRH3 comprises an amino acid sequence having the formula of A<sub>1</sub>-A<sub>2</sub>-(A<sub>3</sub>)<sub>n</sub>-A<sub>4</sub>-A<sub>5</sub>-A<sub>6</sub>-A<sub>7</sub>-A<sub>8</sub>-A<sub>9</sub>; wherein

20 A<sub>1</sub> is an amino acid selected from the group consisting of R, L, and V;

A<sub>2</sub> is an amino acid selected from the group consisting of I, L, and V;

A<sub>3</sub> is any naturally occurring amino acid and n = 1-17;

A<sub>4</sub> is an amino acid selected from the group consisting of E, W, and F;

A<sub>5</sub> is any naturally occurring amino acid;

25 A<sub>6</sub> is an amino acid selected from group consisting of W, G, R, and M;

A<sub>7</sub> is an amino acid selected from the group consisting of V, L, and P; and

A<sub>8</sub> and A<sub>9</sub> are any naturally occurring amino acid.

70. The polypeptide according to claim 40, wherein the polypeptide is a variable  
30 domain of a monobody.

71. The polypeptide according to claim 70, wherein

A<sub>1</sub> is R;

A<sub>2</sub> is I;

A<sub>6</sub> is W;

5 A<sub>7</sub> is V; and

n=9.

72. The polypeptide according to claim 70, wherein

A<sub>1</sub> is L;

10 A<sub>2</sub> is L;

A<sub>4</sub> is W;

A<sub>5</sub> is L; and

n=9.

15 73. The polypeptide according to claim 70, wherein

A<sub>1</sub> is V;

A<sub>2</sub> is L;

A<sub>4</sub> is F;

A<sub>6</sub> is R;

20 A<sub>7</sub> is V; and

n=9.

74. The polypeptide according to claim 70, wherein

A<sub>1</sub> is R;

25 A<sub>2</sub> is L;

A<sub>4</sub> is W; and

n=9.

75. The polypeptide according to claim 70, wherein A<sub>3</sub> is any of the naturally

30 occurring amino acids.

76. The polypeptide according to claim 70, wherein A<sub>3</sub> is encoded by a nonrandom codon set.

77. A polynucleotide molecule encoding a polypeptide of any of claims 59-76.

5

78. A replicable expression vector comprising a polynucleotide molecule of claim 77.

79. A host cell comprising the vector of claim 78.

10 80. A library comprising a plurality of vectors of claim 78, wherein the plurality of vectors encode a plurality of variant polypeptides.

81. A method of generating a polypeptide comprising a variant CDRH3, wherein said polypeptide is capable of binding a target molecule of interest, said method comprising:

15 a) identifying at least one structural amino acid position in CDRH3;

and

b) replacing the amino acid at said at least one structural amino acid position with a variant amino acid found at that position in a population of polypeptides with randomized CDRH3 at a frequency at least one standard deviation above the average frequency for any amino acid at that position; and

20 c) replacing at least one non-structural amino acid position with a variant amino acid, wherein the variant amino acid is any of the naturally occurring amino acid or is encoded by a nonrandom codon set.

25 82. The method according to claim 81, wherein identifying at least one structural amino acid position comprises:

a) generating a population of variant CDRH3 regions from a parent CDRH3 by replacing each amino acid position in the CDRH3 with a scanning amino acid; and

30 b) identifying a structural amino acid position in the CDRH3 as an amino acid position that when substituted with a scanning amino acid, the

substituted polypeptide has a decrease in binding with a target molecule as compared to the parent CDRH3 region, wherein the target molecule specifically binds to a folded polypeptide and does not bind to unfolded polypeptide.

5 83. The method according to claim 81, wherein identifying at least one structural amino acid position comprises:

a) generating a population of polypeptides with randomly generated variant CDRH3 regions, wherein each amino acid position in the variant CDRH3 regions is randomized;

10 b) selecting members of the population that interact with a target molecule, wherein the target molecule specifically binds to a folded polypeptide and does not bind to an unfolded polypeptide;

c) determining the sequence of the selected members; and

15 d) identifying a structural amino acid position as a position that when substituted with a scanning amino acid the substituted polypeptide has a decrease in binding with the target molecule as compared to polypeptide with parent CDRH3 region.

84. The method according to any of claims 81-83, wherein the polypeptide is a

20 variable domain of a camelid monobody.

85. A polypeptide prepared according to the methods of any of claims 81-83.

86. A polynucleotide molecule encoding a polypeptide of claim 85.

25

87. A replicable expression vector comprising a polynucleotide sequence of claim 86.

88. A host cell comprising the vector of claim 87.

30 89. A library comprising a plurality of vectors of claim 87, wherein the plurality of vectors encode a plurality of different polypeptides.

90. A variable domain of a monobody comprising a CDRH3, wherein the CDRH3 comprises an amino acid sequence having the formula of R-A<sub>2</sub>-A<sub>3</sub>-R-(A<sub>5</sub>)<sub>n</sub>; wherein A<sub>2</sub> is L, I or M, A<sub>3</sub> and A<sub>5</sub> are any naturally occurring amino acid, and n = 1 to 20.

5

91. A variable domain of a monobody comprising a CDRH3, wherein the CDRH3 comprises an amino acid sequence having the formula of R-A<sub>2</sub>-(A<sub>3</sub>)<sub>n</sub>-W-A<sub>5</sub>-A<sub>6</sub>-A<sub>7</sub>-A<sub>8</sub>-A<sub>9</sub>; wherein A<sub>3</sub>, A<sub>5</sub>, A<sub>6</sub>, A<sub>7</sub>, A<sub>8</sub> and A<sub>9</sub> are any naturally occurring amino acid and A<sub>2</sub> is L, I or M, and n = 1 to 20.

10

92. A method for designing a CDRH3 scaffold comprising:

- a) generating a library of polypeptides with variant CDRH3 regions;
- b) selecting members of the library that bind to a target molecule that binds to folded polypeptide and does not bind to unfolded polypeptide;
- c) analyzing the binders to identify structural amino acid positions in the CDRH3 region; and
- d) selecting as a scaffold, a binder that has a structural amino acid position at the N and/or C-termini of the CDRH3 and not in a central position of the CDRH3.

20

93. The method according to claim 92, further comprising:

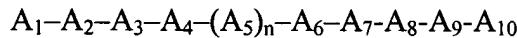
- e) identifying an amino acid that can be substituted at the structural amino acid position, wherein the amino acid is selected from the group of amino acids that occur at that position more frequently than randomly expected;
- f) forming a scaffold with at least one identified amino acid in at least one structural amino acid position.

94. The method according to claim 92 and 93, wherein the structural amino acid positions are selected from the group consisting of the first N-terminal amino acid, the second N-terminal amino acid and the last six C-terminal amino acid, and mixtures thereof.

95. The method according to claim 94, wherein the identified amino acids are selected from the group consisting of arginine, tyrosine, phenylalanine , tryptophan, and valine.

5

96. A polypeptide comprising a CDRH3, wherein the CDRH3 comprises an amino acid sequence having the formula of :



wherein  $A_1$  is an amino acid selected from the group consisting of R, L and V;

10  $A_2$  is an amino acid selected from the group consisting of I, L and V;

$A_3$  is any naturally occurring amino acid;

$A_4$  is selected from the group consisting of C, R and N;

$A_5$  is any naturally occurring amino acid and  $n = 1-16$ ;

$A_6$  is an amino acid selected from the group consisting of C, S, F, T, E and D;

15  $A_7$  is an amino acid selected from the group consisting of W, G, R and M;

$A_8$  is an amino acid selected from the group consisting of V, L and P;

$A_9$  is an amino acid selected from the group consisting of T, V, L and Q; and

$A_{10}$  is an amino acid selected from the group consisting of W, G, S and A.

20 97. The polypeptide according to claim 96, wherein the polypeptide is a variable heavy chain domain.

98. The polypeptide according to claim 97, wherein the polypeptide is a camelid monobody.

25

99. The polypeptide of claim 96, wherein

$A_1$  is R;

$A_2$  is I;

$A_4$  is C;

30  $A_6$  is C;

$A_7$  is W;

A<sub>8</sub> is V;  
A<sub>9</sub> is T;  
A<sub>10</sub> is W; and  
n=7.

5

100. The polypeptide of claim 96, wherein

A<sub>1</sub> is R;  
A<sub>2</sub> is I;  
A<sub>7</sub> is W;  
A<sub>8</sub> is V; and  
n=11.

101. The polypeptide of claim 96, wherein

A<sub>1</sub> is R;  
A<sub>2</sub> is I;  
A<sub>4</sub> is C;  
A<sub>6</sub> is C;  
A<sub>7</sub> is W;  
A<sub>8</sub> is V; and  
n=6.

102. A polynucleotide molecule encoding a polypeptide of any of claims 96-101.

103. A replicable expression vector comprising a polynucleotide molecule of claim

25 102.

104. A library comprising a plurality of vectors of claim 103, wherein the plurality of vectors encode a plurality of variant polypeptides.

30 105. A CDRH3 scaffold comprises a N-terminal portion in which some or all of the positions are structural; and a C terminal portion in which some or all of the amino acid

positions are structural , and wherein the scaffold can accommodate the insertion of a central portion or loop of contiguous amino acids that can vary in sequence and in length.

106. The CDRH3 scaffold of claim 105, wherein the N-terminal portion has a cysteine  
5 residue and the C terminal portion has a cysteine residue, wherein the cysteine residues in the N terminal and C-terminal portion of the CDRH3 scaffold form a disulfide bond that stabilizes the central portion insert, and wherein the central portion insert is a contiguous amino acid sequence of about 1 to 20 amino acids.

10 107. The CDRH3 scaffold of claim 105, wherein the N-terminal portion has a N terminal sequence of R-L/I/M-A<sub>3</sub>-R, wherein A<sub>3</sub> is any naturally occurring amino acid, and wherein the central portion insert is a contiguous amino acid sequence of about 1 to 20 amino acids.

15 108. The CDRH3 scaffold of claim 105, wherein the N terminal sequence is R-I-A<sub>3</sub>-C, wherein A<sub>3</sub> is any naturally occurring amino acid, and wherein the central portion insert is a contiguous amino acid sequence of about 1 to 20 amino acids.

20 109. The CDRH3 scaffold of claim 105, wherein the N terminal sequence comprises R-I, L-L,V-L, or R-L and wherein the central portion insert is a contiguous amino acid sequence of about 1 to 20 amino acids..

25 110. The CDRH3 scaffold of claim 105, wherein the C terminus has a sequence of CWVTW, and wherein the central portion insert is a contiguous amino acid sequence of about 1 to 20 amino acids.

111. The CDRH3 scaffold of claim 105, wherein C-terminal sequence comprises F-X-R-V, W-X-X-L, W-X-M-P, or W-V, wherein X can be any naturally occurring amino acid and wherein the central portion insert is a contiguous amino acid sequence of about 30 1 to 20 amino acids.

112. The CDRH3 scaffold of claim 105, wherein the N terminal portion is about 1 to 4 amino acids.

113. The CDRH3 scaffold of claim 105, wherein the C terminal portion is about 1 to 6  
5 amino acids.

114. The CDRH3 scaffold of claim 105, wherein the central portion is a contiguous sequence of 9 to 12 amino acids.

10